INTERNATIONAL STANDARD

ISO
19005-1

First edition
2005-10-01

Corrected version
2005-12-01

Document management — Electronic document file format for long-term preservation —

Part 1:
Use of PDF 1.4 (PDF/A-1)

Gestion de documents — Format de fichier des documents électroniques pour une conservation à long terme —

Partie 1: Utilisation du PDF 1.4 (PDF/A-1)

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

iTeh STANDARD PREVIEW

(standards.iteh.ai)

ISO 19005-1:2005
https://standards.iteh.ai/catalog/standards/sist/decaef05-7192-422d-84e0-
abe7bd26d3b2/iso-19005-1-2005

Contents

Page

iTeh STANDARD PREVIEW

(standards.itch.ai)

ISO 19005-1:2005
https://standards.iteh.ai/catalog/standards/sist/decaef05-7192-422d-84e0-
abe7bd26d3b2/iso-19005-1-2005

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 19005-1 was prepared by Technical Committee ISO/TC 171, *Document management applications*, Subcommittee SC 2, *Application issues*, in collaboration with Technical Committees ISO/TC 130, *Graphic technology*, ISO/TC 42, *Photography* and ISO/TC 46, *Information and documentation*, Subcommittee SC 11, *Archives/records management*, in a Joint Working Group.

ISO 19005 consists of the following parts, under the general title *Document management — Electronic document file format for long-term preservation*:

— *Part 1: Use of PDF 1.4 (PDF/A-1)*

In this corrected version of ISO 19005-1:2005 paragraph 5 of the Foreword (above) has been augmented in order fo mention the collaboration of ISO Technical Committees ISO/TC 130, ISO/TC 42 and ISO/TC 46 in the preparation of this part of ISO 19005.

Introduction

PDF is a digital format for representing documents. PDF files may be created natively in PDF form, converted from other electronic formats or digitized from paper, microform, or other hard copy format. Businesses, governments, libraries, archives and other institutions and individuals around the world use PDF to represent considerable bodies of important information. Much of this information must be kept for substantial lengths of time; some must be kept permanently. These PDF files must remain useable and accessible across multiple generations of technology. The future use of, and access to, these objects depends upon maintaining their visual appearance as well as their higher-order properties, such as the logical organization of pages, sections, and paragraphs, machine recoverable text stream in natural reading order, and a variety of administrative, preservation and descriptive metadata.

Adobe Systems Incorporated makes the PDF specification publicly available. However, the inclusive, feature-rich nature of the format requires that additional constraints be placed on its use to make it suitable for the long-term preservation of electronic documents.

The primary purpose of this part of ISO 19005 is to define a file format based on PDF, known as PDF/A, which provides a mechanism for representing electronic documents in a manner that preserves their visual appearance over time, independent of the tools and systems used for creating, storing or rendering the files.

A secondary purpose of this part of ISO 19005 is to provide a framework for recording the context and history of electronic documents in metadata within conforming files.

Another purpose of this part of ISO 19005 is to define a framework for representing the logical structure and other semantic information of electronic documents within conforming files.

These goals are accomplished by identifying the set of PDF components that may be used, and restrictions on the form of their use, within conforming PDF/A files.

By itself, PDF/A does not necessarily ensure that the visual appearance of the content accurately reflects any original source material used to create the conforming file; e.g. the process used to create a conforming file might substitute fonts, reflow text, downsample images or use lossy compression. Organizations that need to ensure that a conforming file is an accurate representation of original source material may need to impose additional requirements on the processes that generate the conforming file beyond those imposed by this part of ISO 19005. In addition, it is important for those organizations to implement policies and practices regarding the inspection of conforming files for correct visual appearance.

This part of ISO 19005 should be used as one component of an organization's electronic archival environment for long-term retention of documents. Successful implementation of this part of ISO 19005 for archival purposes depends upon:

— the retention requirements of an organization's archival environment, records management policies and procedures as specified in ISO 15489-1, [9];

— any additional requirements and conditions necessary to ensure the persistence of electronic documents and their characteristics over time, including, but not limited to, those defined by:

    — ISO 14721;

    — ISO/TR 15801, [10];

    — ISO/TR 18492, [12];

    — ISO 18509-1, [13];

    — ISO 18509-2, [14];

— quality assurance processes necessary to verify conformance with applicable requirements and conditions; e.g. an inspection regime to verify the quality and integrity of converted source data.

This part of ISO 19005 should lead to the development of various applications that read, render, write and validate conforming files. Different applications will incorporate various capabilities to prepare, interpret and process conforming files based on needs as perceived by the suppliers of those applications. However, it is important to note that a conforming application must be able to read and process appropriately all files complying with a specified conformance level.

This document has been created as Part 1 of ISO 19005 to allow the creation of future parts, which can provide compatibility with future versions of the underlying PDF specification without rendering this document or applications based on PDF Version 1.4 obsolete.

The following terms, referring to this specification or parts thereof, are recommended when referring to this specification when the full ISO name is not being used:

— "PDF/A" – a synonym for the ISO 19005 family of standards;

— "PDF/A-1" – a synonym for ISO 19005-1;

— "PDF/A-1a" – a synonym for ISO 19005-1 Level A conformance;

— "PDF/A-1b" – a synonym for ISO 19005-1 Level B conformance.

This part of ISO 19005, in conjunction with *PDF Reference* and *XMP Specification*, January 2004, provides sufficient information to interpret any conforming PDF/A file. *PDF Reference* contains a statement from Adobe Systems Incorporated concerning its intellectual property and its willingness to allow perpetual, royalty-free, non-exclusive use of that property in order to promote the use of PDF. Adobe has provided ISO with a similar statement relating to *XMP Specification*. In general, anyone may use *PDF Reference* and *XMP Specification* to create applications that read, write or otherwise process PDF/A files.

Patent claims regarding applications that read, render, write or otherwise process PDF/A files are outside the scope of this part of ISO 19005.

NPES and AIIM (accredited standards developing organizations) maintain an ongoing series of application notes for guiding developers and users of this part of ISO 19005. These application notes are available at <http://www.npes.org/standards/toolspdfa.html> and <http://www.aiim.org/pdfa/app-notes>. Both NPES and AIIM will also retain copies of the specific non-ISO normative references of this part of ISO 19005 which are publicly available electronic documents.

Document management — Electronic document file format for long-term preservation —

Part 1:
Use of PDF 1.4 (PDF/A-1)

1 Scope

This part of ISO 19005 specifies how to use the Portable Document Format (PDF) 1.4 for long-term preservation of electronic documents. It is applicable to documents containing combinations of character, raster and vector data.

This part of ISO 19005 does not apply to:

— specific processes for converting paper or electronic documents to the PDF/A format;

iTeh STANDARD PREVIEW

— specific technical design, user interface, implementation, or operational details of rendering;

(standards.iteh.ai)

— specific physical methods of storing these documents such as media and storage conditions;

ISO 19005-1:2005

— required computer hardware and/or operating systems.ecaef05-7192-422d-84e0-
abe7bd26d3b2/iso-19005-1-2005

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 646, *Information technology — ISO 7-bit coded character set for information interchange*

NOTE 1   The character encoding defined in ISO/IEC 646 is equivalent to ANSI X3.4 (ASCII) [1] and ECMA-6 [2].

ISO/IEC 9541-1, *Information technology — Font information interchange — Part 1: Architecture*

ISO/IEC 10646-1, *Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*

NOTE 2   The character code values defined in ISO/IEC 10646-1 are equivalent to those of Unicode [22].

ISO 14721, *Space data and information transfer systems — Open archival information system — Reference model*

ISO 15930-4, *Graphic technology — Prepress digital data exchange using PDF — Part 4: Complete exchange of CMYK and spot colour printing data using PDF 1.4 (PDF/X-1a)*

Date and Time Formats, W3C Note, 15 September 1997. Available from Internet <http://www.w3.org/TR/NOTE-datetime>

Errata for PDF Reference, third edition, 18 June 2003. Available from Internet <http://partners.adobe.com/asn/acrobat/docs/PDF14errata.txt>

Extensible Markup Language (XML) 1.0 (Third Edition), W3C Recommendation, 4 February 2004. Available from Internet <http://www.w3.org/TR/2004/REC-xml-20040204>

ICC.1:1998-09, *File Format for Color Profiles*, International Color Consortium. Available from Internet <http://www.color.org/ICC-1_1998-09.PDF>

ICC.1A:1999-04, *Addendum 2 to Spec. ICC.1:1998-09*, International Color Consortium. Available from Internet <http://www.color.org/ICC-1A_1999-04.PDF>

PDF Reference: Adobe Portable Document Format, Version 1.4, Adobe Systems Incorporated – 3rd ed. (ISBN 0-201-75839-3). Available from Internet <http://partners.adobe.com/asn/acrobat/docs/File_Format_Specifications/PDFReference.pdf>

RDF/XML Syntax Specification (Revised), W3C Recommendation, 10 February 2004. Available from Internet <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>

Tags for the Identification of Languages, RFC 1766, March 1995. Available from Internet <http://www.ietf.org/rfc/rfc1766.txt>

XMP Specification, January 2004, Adobe Systems Incorporated. Available from Internet <http://partners.adobe.com/asn/tech/xmp/pdf/xmpspecification.pdf>

NOTE 3   AIIM and NPES (accredited standards developing organizations) maintain copies of the non-ISO references that are publicly available electronic documents.

3   Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1
conformance level
identified set of restrictions and requirements to which files and readers must comply

[ISO 15930-4]

3.2
cross reference table
PDF data structure that contains the byte offset of the start of indirect objects within the file

3.3
dictionary
associative table containing key-value pairs, specifying the name and value of an attribute for objects, which is generally used to collect and tie together the attributes of a complex object

[ISO 15930-4]

3.4
electronic document
electronic representation of a page-oriented aggregation of text and graphic data, and metadata useful to identify, understand and render that data, that can be reproduced on paper or optical microform without significant loss of its information content

3.5
end-of-file marker
five character sequence **%%EOF** marking the end of a PDF file

3.6
end-of-line marker
EOL marker
one or two character sequence marking the end of a line of text, consisting of a **CARRIAGE RETURN** character (0Dh) or a **LINE FEED** character (0Ah) or a **CARRIAGE RETURN** followed immediately by a **LINE FEED**

3.7
font
identified collection of graphics that may be glyphs or other graphic elements

[ISO 15930-4]

3.8
glyph
recognizable abstract graphic symbol that is independent of any specific design

[ISO/IEC 9541-1]

3.9
ICC profile
colour profile conforming to the ICC specification and its addendum

[ICC.1:1998-09] and [ICC.1A:1999-04]

3.10
interactive reader
reader that requires or allows human interaction during the software's processing phase

NOTE   A file viewing tool is an example of an interactive reader; a raster image processor is an example of a reader that is not interactive.

3.11
Level A conformance
conformance level encompassing all requirements of this part of ISO 19005

3.12
Level B conformance
conformance level encompassing the requirements of this part of ISO 19005 regarding the visual appearance of electronic documents, but not their structural or semantic properties

3.13
long-term
period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository, which may extend into the indefinite future

[ISO 14721]

3.14
PDF
Portable Document Format
file format defined in *PDF Reference* and its *Errata*

[ISO 15930-4]

3.15
reader
software application that is able to read and process files appropriately

[ISO 15930-4]

3.16
space character
text string character used to represent orthographic white space in the operands of text-showing operators

NOTE   Commonly used space characters include **HORIZONTAL TABULATION** (U+0009), **LINE FEED** (U+000A), **VERTICAL TABULATION** (U+000B), **FORM FEED** (U+000C), **CARRIAGE RETURN** (U+000D), **SPACE** (U+0020), **NO-BREAK SPACE** (U+00A0), **EN SPACE** (U+2002), **EM SPACE** (U+2003), **FIGURE SPACE** (U+2007), **PUNCTUATION SPACE** (U+2008), **THIN SPACE** (U+2009), **HAIR SPACE** (U+200A), **ZERO WIDTH SPACE** (U+200B), and **IDEOGRAPHIC SPACE** (U+3000).

3.17
white-space character
NULL (00h), **HORIZONTAL TABULATION** (09h), **LINE FEED** (0Ah), **FORM FEED** (0Ch), **CARRIAGE RETURN** (0Dh) or **SPACE** (20h) character

3.18
writer
software application that is able to write files

[ISO 15930-4]

3.19
XMP packet
structured wrapper for serialized XMP metadata that can be embedded in a wide variety of file formats

4   Notation

PDF operators, PDF keywords, the names of keys in PDF dictionaries, and other predefined names are written in bold sans serif font; operands of PDF operators or values of dictionary keys are written in italic sans serif font.

EXAMPLE   The *Default* value for the **TR2** key.

Token characters used to delimit objects and describe the structure of PDF files, as defined in *PDF Reference* 3.1, may be identified by their ISO/IEC 646 character name written in upper case in bold sans serif font followed by a parenthetic two digit hexadecimal character value with the suffix "h".

EXAMPLE   **CARRIAGE RETURN** (0Dh).

Text string characters in content streams, as defined by *PDF Reference* 3.8.1, may be identified by their ISO/IEC 10646-1 character name written in uppercase in bold sans serif font followed by a parenthetic four digit hexadecimal character code value with the prefix "U+".

EXAMPLE   **EN SPACE** (U+2002).

For the purposes of this part of ISO 19005, references to the "*PDF Reference*" are to *PDF Reference: Adobe Portable Document Format*, version 1.4, 3rd ed., as amended by *Errata for PDF Reference*, 3rd ed.

5   Conformance levels

5.1   General

This part of ISO 19005 defines a file format for representing electronic documents known as "PDF/A-1." Conforming PDF/A-1 files shall adhere to all requirements of *PDF Reference* as modified by this part of ISO 19005. A conforming file may include any valid *PDF Reference* feature that is not explicitly forbidden by this part of ISO 19005. Features described in PDF specifications prior to Version 1.4 which are not explicitly described in *PDF Reference* should not be used. Neither the version number in the header of a PDF file nor

the value of the **Version** key in the document catalog dictionary shall be used in determining whether a file is in accordance with this part of ISO 19005.

NOTE 1   A conforming file is not obligated to use any PDF feature other than those explicitly required by *PDF Reference* or this part of ISO 19005.

NOTE 2   The proper mechanism by which a file can presumptively identify itself as being a PDF/A-1 file of a given conformance level is described in 6.7.11.

5.2   Level A conformance

Level A conforming files shall adhere to all of the requirements of this part of ISO 19005. A file meeting this conformance level is said to be a "conforming PDF/A -1a file."

5.3   Level B conformance

In recognition of the varying preservation needs of the diverse user communities making use of PDF files, this part of ISO 19005 defines a Level B conformance level. Level B conforming files shall adhere to all of the requirements of this part of ISO 19005 except those of 6.3.8 and 6.8. A file meeting this conformance level is said to be a "conforming PDF/A-1b file."

NOTE   The Level B conformance requirements are intended to be those minimally necessary to ensure that the rendered visual appearance of a conforming file is preservable over the long term. However, Level B conforming files might not have sufficiently rich internal information to allow for the preservation of the document's logical structure and content text stream in natural reading order, which is provided by Level A conformance. The requirements for Level A conformance place greater responsibilities on writers of conforming files and those preparing such files, but these requirements allow for a higher level of document preservation service and confidence over time. Additionally, Level A conformance facilitates the accessibility of conforming files for physically impaired users.

5.4   Conforming readers

A conforming reader shall comply with all requirements regarding reader functional behaviour specified in this part of ISO 19005. The requirements of this part of ISO 19005 with respect to reader behaviour are stated in terms of general functional requirements applicable to all conforming readers. This part of ISO 19005 does not prescribe any specific technical design, user interface or implementation details of conforming readers.

The rendering of conforming files shall be performed as defined in *PDF Reference* subject to the further requirements specified by this part of ISO 19005. Features described in PDF specifications prior to Version 1.4 that are not explicitly described in *PDF Reference* may be ignored by conforming readers.

Conforming readers shall read and process appropriately all PDF/A-1 files complying with a specified conformance level. Level A conforming readers shall read and process appropriately all Level A and B conforming files. Level B conforming readers shall read and process appropriately all Level B conforming files.

6   Technical requirements

6.1   File structure

6.1.1   General

6.1.2 to 6.1.13 address overall file format issues and the base elements that form the general structure of a conforming file.

6.1.2   File header

The **%** character of the file header shall occur at byte offset 0 of the file.

The file header line shall be immediately followed by a comment consisting of a **%** character followed by at least four characters, each of whose encoded byte values shall have a decimal value greater than 127.

NOTE   The presence of encoded character byte values greater than decimal 127 near the beginning of a file is used by various software tools and protocols to classify the file as containing 8-bit binary data that should be preserved during processing.

6.1.3   File trailer

The file trailer dictionary shall contain the **ID** keyword. The keyword **Encrypt** shall not be used in the trailer dictionary. No data shall follow the last end-of-file marker except a single optional end-of-line marker.

The file trailer referred to is either the last trailer dictionary in a PDF file, as described in *PDF Reference* 3.4.4 and 3.4.5, or the first page trailer in a linearized PDF file, as described in *PDF Reference* F.2. In a linearized file the **ID** keyword shall be present in both the first page trailer and the last trailer dictionaries and the value of both instances of the keyword shall be identical.

NOTE   The explicit prohibition of the **Encrypt** keyword has the implicit effect of disallowing encryption and password-protected access permissions.

6.1.4   Cross reference table

In a cross reference subsection header the starting object number and the range shall be separated by a single **SPACE** character (20h).

The **xref** keyword and the cross reference subsection header shall be separated by a single EOL marker.

Any object whose offset is not referenced in the cross reference table shall be exempt from all requirements of this part of ISO 19005.

6.1.5   Document information dictionary

A document information dictionary may be defined in a conforming file. If defined, its elements shall be consistent with analogous XMP metadata properties as specified in 6.7.3.

6.1.6   String objects

Hexadecimal strings shall contain an even number of non-white-space characters, each in the range **0** to **9**, **A** to **F** or **a** to **f**.

6.1.7   Stream objects

The **stream** keyword shall be followed either by a **CARRIAGE RETURN** (0Dh) and **LINE FEED** (0Ah) character sequence or by a single **LINE FEED** character. The **endstream** keyword shall be preceded by an EOL marker.

The value of the **Length** key specified in the stream dictionary shall match the number of bytes in the file following the **LINE FEED** character after the **stream** keyword and preceding the EOL marker before the **endstream** keyword.

NOTE 1   These requirements remove potential ambiguity regarding the ending of stream content.

A stream object dictionary shall not contain the **F**, **FFilter**, or **FDecodeParams** keys.

NOTE 2   These keys are used to point to document content external to the file. The explicit prohibition of these keys has the implicit effect of disallowing external content that can create external dependencies and complicate preservation efforts.